An automatic method to create flow lines for determination of glacier length: A pilot study with Alaskan glaciers

Le Bris, Raymond ; Paul, Frank

Abstract: Glacier length is a key parameter in global glacier inventories, but difficult to determine in a consistent way and subject to frequent change. Its vector representation (a flow line) is a most important input for modeling future glacier evolution, but only seldom available from digital databases. Hence, there is an urgent need to generate such flow lines for a large number of glaciers from automated methods. We here present a new algorithm that is based on Python scripting and additional libraries (GDAL and OGR) and requires only a DEM and glacier outlines as an input. The core of the method is based on a glacier axis concept that is combined with geometry rules such as the k-d Tree, Nearest Neighbor and crossing test theory. We have applied the method to 400 glaciers located in Western Alaska, where a new glacier inventory was recently created. The accuracy of the method was assessed by a quantitative and qualitative (outline overlay) comparison with a manually digitized dataset for 20 glaciers. This comparison revealed for 17 out of the 20 glaciers a length value within the range of the manual digitizations. Other potential methods performed less well. Combined with previous glacier outlines from the same region (Digital Line Graph) we automatically determined length changes for 390 glaciers over a c. 50 year period.

DOI: https://doi.org/10.1016/j.cageo.2012.10.014

Corresponding author:
Raymond Le Bris
E-mail: raymond.lebris@geo.uzh.ch
Telephone: +41 - 44 - 6355208
Fax: +41 - 44 - 6356841

An automatic method to create flow lines for determination of glacier length: A pilot study with Alaskan glaciers

Raymond Le Bris, Frank Paul
Department of Geography, University of Zurich,
Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
E-mail: raymond.lebris@geo.uzh.ch

Computers & Geosciences (in press)

ABSTRACT

Glacier length is a key parameter in global glacier inventories, but difficult to determine in a consistent way and subject to frequent change. Its vector representation (a flow line) is a most important input for modelling future glacier evolution, but only seldom available from digital databases. Hence, there is an urgent need to generate such flow lines for a large number of glaciers from automated methods. We here present a new algorithm that is based on Python scripting and additional libraries (GDAL and OGR) and requires only a DEM and glacier outlines as an input. The core of the method is based on a glacier axis concept that is combined with geometry rules such as the k-d Tree, Nearest Neighbour and crossing test theory. We have applied the method to 400 glaciers located in Western Alaska, where a new glacier inventory was recently created. The accuracy of the method was assessed by a quantitative and qualitative (outline overlay) comparison with a manually digitized dataset for 20 glaciers. This comparison revealed for 17 out of the 20 glaciers a length value within the range of the manual digitizations. Other potential methods performed less well. Combined with previous glacier outlines from the same region (Digital Line Graph) we automatically determined length changes for 390 glaciers over a c. 50 year period.

Keywords: Glacier; Remote sensing; Flow lines; Algorithm

1. INTRODUCTION

Glaciers are regarded as natural elements documenting climate change most clearly to a wide public (Lemke et al., 2007). For this and further reasons (e.g. their high sensitivity to climate change) glaciers were considered as one of the terrestrial essential climate variables (ECVs) by the Global Climate Observing System (GCOS 2003). In the last century, glaciers worldwide experienced a strong decline (retreat and mass loss) with only a few local exceptions (WGMS 2008). Whereas mass changes of a glacier are a direct and undelayed response to the atmospheric conditions in the respective year, changes in glacier length (or terminus fluctuations) are a delayed, filtered and enhanced response to atmospheric conditions over a climatically relevant period of a few decades. Though the link of glacier terminus fluctuations to climatic forcing is difficult to establish, it can be made nevertheless (Hoelzle et al., 2003; Klok & Oerlemans, 2004). The special advantages of terminus fluctuations are: (1) their long historical record, partly back to the 16th century

(Nussbaumer & Zumbühl, 2011), and (2) their widespread availability from numerous mountain ranges. Length changes are thus a key element for the reconstruction of past climatic fluctuations (Oerlemans, 2005; Leclercq & Oerlemans, 2011) or changes in sea level (Oerlemans et al., 2007, updated by Leclercq et al., 2011).

However, the sample of glaciers that can be used in such assessments is small compared to the estimated total number of glaciers on Earth (about 200,000: Arendt et al., 2012) and might be considered as being biased (e.g. in regard to their size or location). This sample can be extended by using multitemporal satellite data (e.g. Paul et al., 2011) that document glacier changes over recent decades in all parts of the world. The satellite data can also be used to map the Little Ice Age (LIA) extent from trimlines and thus largely extend the time series (Lopez et al., 2010; Citterio et al., 2010; Glasser et al., 2011). As the length changes since that time are often measured in kilometres, the low spatial resolution of sensors such as Landsat has only a small effect on the quality of the results (Hall et al., 2003). One bottleneck for a wider application of such satellite-based length change measurements is the work load involved: as yet, the points for measuring changes have to be defined and digitized manually. To largely extend this application, there is a demand to find an appropriate reference point at the glacier terminus automatically. Though the lowest elevation of a glacier can be determined automatically within a Geographic Information System (GIS) from glacier outlines and a digital elevation model (DEM), this is only meaningful for glaciers with a sharp tongue and/or in steep terrain. For glaciers with a flat or wide tongue such a point can be at almost any location on the glacier front and it would be beneficial to define such a point near the centre of the terminus. This can be achieved by digitally intersecting a flow line that is located in the centre of a glacier tongue (in the ablation region) with the glacier outline at the terminus (Paul et al., 2009). Hence the problem of automatically calculating terminus fluctuations also implies the need for automated creation of a flow line.

As mentioned above, glacier length has a key role in glacier inventories: As a scalar value length is a major input for several modelling approaches (Haeberli & Hoelzle, 1995; Bahr, 1997; Lüthi et al., 2010); in its vector form the flow line is a mandatory input for flow models that assess future glacier changes (e.g. Oerlemans, 2008). So both are needed, the scalar value and the vector line itself. They are, however, currently not part of most glacier data stored in the GLIMS database (Raup et al., 2007). As manual digitizing does not provide reproducible or consistent results, an automated determination would be highly preferable. This is particular true in times of strong geometric glacier changes (Paul et al., 2007), as in such cases the entire flow line needs to be digitized again for a new inventory. An automated determination will reduce the amount of work (and hence increase the number of glaciers for which a flow line will be available) and give reproducible results. With flow line we refer here to the terminology introduced for glacier inventories decades ago to describe glacier length (e.g. Müller et al., 1977). We thus do not refer to a particle trajectory in a strict physical sense, but are compliant with the flow lines practically used for glacier modelling, i.e. being located in the centre of a glacier tongue in the ablation region (cf. Oerlemans, 2001). The method of calculation does not result in a longest flow line or a line providing the mean glacier length from several branches as defined in Müller et al., (1977). It is just seen as a line providing a reasonable and reproducible value and vector segment for glacier length.

Besides the flow line, the automatic algorithm presented here creates points on the glacier terminus using only glacier outlines and a DEM as an input. The algorithm is thus widely applicable and avoids the shortcomings of approaches that require strong intervention by the analyst. The method is applied to a test region in Alaska, where glaciers have a large variety of shapes and a new glacier

inventory was recently created from satellite data (Le Bris et al., 2011). For validation we performed a comparison of the algorithm results with manually created flow lines. The quality of the generated dataset is further demonstrated by a comparison of its results with two other approaches that are also applicable to large datasets. The flow lines are finally used to determine length changes for a sample of about 400 glaciers by digital intersection with earlier outlines that are available for this region from the US Geological Survey (USGS).

2. INPUT DATA AND TEST REGION

2.1 Study region

While the outline dataset from the new glacier inventory (Le Bris et al., 2011) covers a large proportion of all Alaskan glaciers, we focus in this study on the western part of it. The study region is located in the southern part of the Alaska Range and includes the Tordrillo and the Chigmit Mountains (Fig. 1) with glaciers covering ca. 2000 and 2800 km^2, respectively. The Gulkana and Wolverine glaciers are located close to the study region (at 63°16'N, 145°25'E and 60°24'N, 148°54'E, respectively) and are separately analysed as they are benchmark glaciers in terms of long-term mass balance observations (WGMS, 2009). Glaciers can be found in the study region at all altitudes ranging from sea level up to 4000 m a.s.l. Several glacier types are present with large valley glaciers of complex shape (like the 213 km^2 Trimble glacier), but also small cirque glaciers and glaciers on volcanoes (Denton & Field, 1975). The climate regime is predominantly of maritime type close to the coast, but gets more continental further inland (e.g. http://climate.gi.alaska.edu).

2.2 Datasets

Two kinds of input data are needed to apply the flow line algorithm: A DEM with sufficient quality (e.g. not too many artefacts, local sinks and data voids) and glacier outlines in a vector format. If the geomorphometric representation of the glacier surface is too poor in the DEM, some filtering and editing is required beforehand. In this study we used the first release of the ASTER GDEM rather than the Shuttle Radar Topography Mission (SRTM) DEM because most of the study region (apart from South Kenai Peninsula) is located north of the 60° North limit of that DEM. The ASTER GDEM was compiled from all available scenes in the ASTER archive acquired between 2000 and 2007. For change assessment, we additionally used a subset of glacier outlines from the new Alaskan glacier inventory referring to the year 2007 (+/-2 years), the USGS National Elevation Dataset (NED) (Gesch, 2007), the Digital Line Graph (DLG) and the Digital Raster Graph (DRG), which all refer to a former glacier surface topography and extent. The USGS 1:63,360-scale 15-minute topographic quadrangle maps were created from vertical aerial photographs (acquired between 1948 and1957) by stereophotogrammetric techniques. These maps were scanned (http://topomaps.usgs.gov/drg) and used to both create the NED DEM by extracting contour lines and to compile the DLG glacier outlines (http://eros.usgs.gov/#/Guides/dlg). All datasets are re-projected to UTM zone 5 with WGS84 datum and 30 m cell size.

2.3 Pre-processing of the raw data

The satellite-derived glacier outlines were manually corrected for classification errors during post-processing. This concerns in particular debris-covered glacier parts, shadow regions and attached seasonal snow fields (e.g. Racoviteanu et al., 2009). For most glaciers these outlines and related terminus positions are assumed to be accurate to about +/-1 pixel (30 m) or +/-2 pixels in the case of debris cover on the glacier and poor contrast to the glacier forefield.

The drainage divides that are required to clip the contiguous ice masses into individual glaciers were derived from hydrological calculations within a GIS using the NED DEM and following a method described by Bolch et al., (2010). These drainage divides were also applied to the DLG outlines to refer to the same glacier entities in both datasets. Hillshades were also created from the DEM to aid in visual interpretation during the manual digitizing experiment.

As the glacier outlines in the DLG dataset were partly updated, the DLG extents were adjusted to the DRG maps to have a clear reference date for each glacier extent. Unfortunately, the DRG maps showed a systematic planimetric shift compared to the DLG that might have resulted from different parameters and/or software used during the georeferencing process. Hence, we first have corrected this displacement by applying a correction to the coordinates of each raster file (60 to 90 m in x and 90 to 150 m in y). This helped to achieve a very good match of the USGS vector dataset to the raster maps. The differences between the DLG and the DRG were mainly found in the ablation areas and were manually corrected (using the shifted DRG in the background) for a subset of 400 glaciers larger than 1 km^2. This dataset was later used for change assessment and covers a larger region than the study area selected for the methodological development (see Fig. 1).

Fig. 1

3. METHODS

3.1 Background

Glaciers originate from the accumulation of snow that is slowly transformed into ice by metamorphism and finally flows downward under its own weight. The resulting size and shape of a glacier does strongly depend on the climatic conditions and the topographic characteristics of the respective region. A glacier can thus have any geometry, ranging from a very simple elongated valley glacier (e.g. North Twin Glacier) to a more complex topology with many large multi-tributary branches (e.g. Triumvirate Glacier). In this regard, a suitable approach should be sufficiently flexible to create a flow line for any glacier form. On the other hand, glacier length as a parameter is highly ill-defined for complex shapes. Even when length is more precisely defined as the longest flow line or the line connecting the highest with the lowest point of a glacier, this does not necessarily result in a unique assignment for all glaciers as the path in-between is still subject to different interpretations. There is thus in any case some variability in this parameter that has to be considered when comparing a modelled value to a 'ground truth' (i.e. a manually digitized line).

The flow lines created here have to fulfil several rules to provide glacier length: The line must be located inside a glacier outline; basically stretch from the highest to the lowest point, should not cross rock outcrops, and should always be in the centre of a glacier, in particular in the ablation region. However, considering that emerging rock outcrops are not part of the glacier, the line from the algorithm will correctly curves around them. Also in this case the line will not be exactly located in the centre of a glacier. Whereas the human brain is able to consider all these points during manual digitization (though not quite reproducible), this is much more difficult for an automated algorithm that can only consider a certain number of rules (one after another). And 'curving around a rock outcrop' is actually a highly complex concept, even in robotics using artificial intelligence. Hence, to get the method performing in a reasonable amount of time, we apply simple rules that are flexible enough to get a large number of glacier shapes correctly processed. Indeed, the algorithm will not work correctly for all cases and we can expect some post-processing work to correct misplaced flow lines. For the two glacier inventories (DLG and satellite-derived), distinct polyline

(flow lines) and point (terminus position) shapefiles were generated by the algorithm which took about two hours of processing time. This time depends on the computer system used to perform the algorithm and could be either slightly faster or slower with an up-to-date or old systems. The individual steps of the algorithm are described in section 4.3.

3.2 Manual digitizing experiment

To assess the performance of the automated flow line delineation, a manual digitization experiment was performed. It is based on the repeat digitization of flow lines from 20 glaciers (covering sizes from 1 km^2 to ca. 250 km^2) by three persons, three times each. Each digitization was done with at least one day in-between to avoid recognition of the previous positioning of the line. False colour composite images, the DEM hillshade and 50 m contour lines derived from the DEM were used to aid in the visual interpretation. The start and end points were given, but neither the automatically derived flow lines nor the results of the other participants were available. This procedure guarantees at least a certain independence of the individual digitizations. The dataset allows three comparisons to be performed: (a) variability of the manual digitization in absolute terms, (b) determination of the difference to the automatically derived length values, and (c) overlay of all lines to identify the problematic regions.

3.3 Other options to determine glacier length

Previous to the development of this algorithm, we performed tests with other approaches, for example using a water routing algorithm starting from seed points (e.g. Quinn et al., 1991). Although these tests gave good results in the accumulation region (with its concave shape), this method fails in the ablation area due to the convex shape of glacier tongues: they leave the glacier and can thus not stop at the terminus. Also in the accumulation region they are not really located in the centre but follow the steepest downward gradient in the DEM, making them highly susceptible for DEM artefacts. Hence, they neither provide a suitable length value nor can they be used for assessment of length changes.

There are further possibilities to derive at least a scalar value for glacier length from easily available input data. Though this has clear shortcomings compared to direct determination from a flow line, they can be computed very fast for a large number of glaciers and we find a comparison interesting in regard to the potential error bounds of the automated method. One method is based on the high correlation between glacier area and length that is obviously based on a self-correlation (width multiplied with length estimates area). We have calculated a corresponding regression between the modelled length values and the area from the glacier inventory and applied this relation to the 20 selected test glaciers. The second method used for comparison is based on an inversion of the calculation of mean slope from the arctangent of glacier elevation range and length. We here use mean slope and elevation range as calculated per glacier from the DEM and outlines (using the ArcGIS zonal statistic tools) and calculate glacier length L from: L = elevation range / tan (slope).
This should work pretty well for constantly sloped (mountain) glaciers, but will likely underestimate real glacier length for long valley glaciers with a curved profile.

3.4 Calculating length changes

Using the adjusted glacier extents from the DLG, the outlines from the satellite-derived glacier inventory, and the flow lines derived here, we calculated length changes by digital intersection of the flow line for the DLG extent with the more recent outlines. Strictly speaking, this is only the change of the terminus position (or front variation) rather than the change of the entire length, as the latter can also result from a new path a flow line has to take due to changes in geometry (e.g.

emerging rock outcrops). From a glacier dynamical and modelling point of view, only the frontal variation is of interest. We also computed the Euclidian distance between terminus positions of each glacier from the two inventories to assess the efficiency of this method. In total, 400 glaciers larger than 1km^2 in both datasets served as a test sample.

4. IMPLEMENTATION

4.1 GDAL-OGR Library and Python language

In order to create a cross-platform and/or software independent tool, the algorithm is written in Python. Python is an open source programming language with an efficient high-level data structure that uses functions and data types implemented in C or C++ (http://www.python.org). As we have to handle different sorts of data, we have implemented an additional library in the Python environment that allows us to properly work with glacier outlines and a DEM. In this regard, the GDAL-OGR (Geospatial Data Abstraction Library – Open source Geospatial Resources) library is used as a translator between raster and vector geospatial data formats (http://www.gdal.org).

4.2 Finding the highest glacier point

As described in the next section, the algorithm derives flow lines by joining the highest elevation point of any given glacier to its respective terminus position through several other middle points. The first task is thus to identify the highest point for each glacier, at best automatically. We have tested two different approaches to get these points called hereafter ‘starting points’. The first one (A) determines the starting point as the highest elevation inside each glacier outline. The main advantage of this approach is that those points are derived automatically for several hundred glaciers at once using GIS tools. The disadvantage is that the lines obtained starting at these points will not necessarily be the longest. Indeed, when the purpose is also to assess length changes, one has to keep in mind that the points could be outside the outline of the second glacier (from a newer or older date) with the consequence of reducing the number of comparable glaciers. The second approach (B) is to manually create the starting points with two possibilities: (1) first, automatically compute the highest points and then manually move them to a slightly displaced location to get them common to both datasets, or (2) directly place the points in what appears to be the longest branch. Choosing between (1) and (2) will mainly depend on the number of glacier in the dataset. In the case of a small dataset we recommend manual creation of the points (2), otherwise option (1) is the more practical.

4.3 Work flow of the algorithm

The algorithm described here creates a flow line for a given glacier and applies the individual steps depicted in Fig. 2 in a loop to an entire set of glaciers. Considering that a flow line is simply a line, it can be seen by definition as a succession of points that are linked to each other. The core principle of the algorithm is thus to find middle points inside any given glacier outline at all altitude bands and link them together. Though it is straightforward to find these points for simple polygon geometry, this is different for glaciers that could have any shape. To solve this problem, we have developed the algorithm based on what we call the glacier axis concept. The individual steps of the work flow are explained in the following (cf. Fig. 2):

a) Compute the starting and end points: Using glacier outlines, a DEM and the zonal statistics approach, the highest and lowest points for each glacier are calculated and stored in its attribute table. Assuming that a given glacier can have multiple points with the same elevation, only the first point in the resulting list is selected. Manual corrections are applied where necessary.

b) Create the glacier axis: As the base for the axis concept we assume that the main direction of any given glacier can be defined as a straight line from its highest to its lowest elevation. In a first approach, this is assumed to give the general direction of flow and is further used to derive traverses across each glacier (c). In some specific cases (i.e. the glacier axis does not characterize the maximum elongation of the glacier geometry) the glacier axis needs to be extended. This is achieved by extending the line by 25% at both ends for all glaciers.

c) Create perpendicular traverses: The algorithm then computes traverses at regular horizontal intervals along the axis (in 100 m steps). These traverses are used to slice the glacier by intersecting the glacier outline. In order to entirely cover the glacier, all traverses are set by default to 50 km width (25 km on each side) as no glacier in the new Alaskan glacier inventory is wider than 49 km (Columbia Glacier). For this step it has to be noted that the elevation intervals between traverses will affect the number of middle points and thus the smoothness of the resulting flow line. For a higher number of middle points there are more possibilities for the algorithm to find the way towards the end point of the given glacier.

d) Intersect traverses with glacier outline: Using OGR library tools (see 4.1), the algorithm intersects traverses and glacier outlines. All traverses and glaciers are linked by a common ID.

e) Compute the middle point of each segment: Because a traverse can be split into many segments from the glacier shape, the middle point of every segment is computed.

f) Get the elevation value for each middle point: Elevations of middle points derived in step e) are extracted from the DEM. Then, the algorithm will link all points following the rules in (g).

g) Create the flow line: After middle points have been computed (step e), the Flow Line Algorithm (hereafter called FLA) will link them to create the flow line following the rules listed below. This is accomplished by implementing some computing science concepts in the script such as: (1) the k-Dimensional Tree (kd-tree), (2) Nearest Neighbour (NN), and (3) crossing test theories. A kd-tree is a data structure for storing a finite set of points from a k-dimensional space (Moore, 1991). This space-partitioning technique efficiently increases the processing speed in searching for the nearest neighbour point to a given position by answering a query for a large set of points (Manolopoulos et al., 2005; Tsaparas & Science, 1999). Concept (3) refers to a line segment intersection theory used in the FLA to test whether or not the nearest point will be selected as being a part of the flow line (De Berg et al., 2008). Rules that have to be followed here are:

(1) Go downward from the starting to the end points,
(2) Do not cross glacier outlines (including internal rock outcrops),
(3) Go from each point to its nearest neighbour.

While (1) implies that in most cases the flow lines are not the longest ones (i.e. for a glacier with several tributaries), (3) constrains the algorithm to link the current point with a lower nearest neighbour point in order to approximate the natural glacier flow. This constrain also prevents for possible altitude inversion in DEMs especially in the ablation regions. It has to be noted here, that DEM accuracy will also affect the automated creation of flow lines. For that reason, we suggest smoothing the DEM with a low-pass filter (e.g. 3x3 median filter) to remove outliers and fill local sinks.

h) Smooth the flow line: This last step is given as a suggestion for the user and is not performed by the algorithm. Its intention is to smooth the lines to give them a more realistic shape. In this study we applied the Kernel (PAEK) algorithm (Bodansky et al., 2002) using ArcGIS software for this purpose. Users can apply different parameters to this polynomial approximation depending on the glacier shape and the required smoothing of the automatically generated flow lines.

Fig. 2

5. RESULTS

5.1 Glacier length

With the FLA we have computed glacier length values for a total of 788 entities from the DLG and the new glacier inventory. Vector lines are obtained for 98% of the processed glaciers (393) for the DLG and 98.7% (395) for the new inventory subset. In some cases the algorithm failed to create flow lines because of a particular glacier shape, resulting in 388 common lines (98.4%). However, due to strong changes in glacier topography and/or geometry, many lines are not directly comparable (see 5.2). This implies that length changes (i.e. front variations) cannot be obtained by direct subtraction of the length values of both lines.

5.2 Visual comparison of the flow lines

Figure 3 shows an overlay of the flow lines obtained with the FLA from (a) the outlines of the new glacier inventory combined with the ASTER GDEM and (b) the DLG outlines with the NED DEM for the two glaciers Wolverine and Gulkana. In the case of Gulkana glacier, the two flow lines differ in shape due to the emerging rock outcrops at the confluence with the tributary, which results in a deflection of the flow direction. As there are only minor changes in the outline of Wolverine glacier but strong changes in surface elevation between the two DEMs, the differences in position are here due to the change of the DEM.

Fig. 3

The overlay of the flow lines from the digitizing experiment (for glaciers #9 and #17) revealed only small overall differences in the positions from one session or operator to another (Fig. 4). However for some glaciers and more locally, also larger variations occurred, emphasizing the difficulty in digitizing all parts of the line consistently. Overall, the line derived with the FLA is more or less in the centre of the manual digitizations pointing to its usefulness for other applications. However, the highest and lowest elevation points have to be provided to the analyst otherwise the digitizations could differ significantly (mainly due to a different consideration of tributaries). In any case, the line from the FLA has the clear benefit of being a consistent and reproducible product which is particularly useful for a glacier inventory.

Fig. 4

5.3 Validation

A scatter plot of length values for all manually digitized flow lines compared to the FLA result is depicted in Fig. 5. For 17 out of 20 lines, the lengths of the FLA lines are within the variability of the lines that had been manually digitized. In three cases (#8, #15 and #16, Fig. 6) the automated lines are outside this variability. Glacier #8 (Blockade glacier) is a large glacier (256 km^2) with a

complex shape encompassing many tributaries and rock outcrops (mainly in the accumulation area) with a tongue flowing down in two opposite sides of a flat valley (Fig. 6A). These glacier characteristics can explain the spurious shape of the automated line which obviously needs to be corrected. For glaciers #15 and #16 the interpretation of the difference is less straightforward since the glacier shapes are rather usual (Fig. 6B and 6C). We speculate that DEM artefacts might explain these outliers and conclude that a visual control of the results and potential correction is required in the post-processing step.

Fig. 5
Fig. 6

The FLA length values as obtained for the new inventory were at first correlated with glacier area to investigate whether area can be used as a predictor for length. Figure 7a shows the related log-log plot of length vs. area yielding a correlation coefficient of $R^2 = 0.88$. This seems to be a good correlation at first glance, but when plotting the relative differences versus glacier length (Fig. 7b), in particular shorter glaciers show large deviations of 50% or more compared to the length [L] computed from the algorithm. This might result from the much larger variability in width compared to the length for such small glaciers.

Fig. 7

The comparison of glacier length as derived from mean glacier slope [α] and elevation range [Dh], ($L_{slope}=\Delta h / \tan \alpha$) is depicted in Fig. 8a. The correlation coefficient $R^2 = 0.83$ is somewhat lower here and compared to the identity the length is systematically underestimated. This is expected, as a curved line between two points is always longer than the direct connection. When plotting the relative difference of [L_{slope}] and [L] vs [L] (Fig. 8b) a similar strong scatter as seen in Fig. 7b is obvious for small glaciers, but the range of the differences is smaller and the differences increase towards longer glaciers. This is in contrast using area to derive length, where differences are getting smaller for larger glaciers (though with a larger scatter). However, despite these interesting differences neither area nor slope is sufficiently accurate predictors for glacier length when calculated from a regression.

Fig. 8

Lengths as computed from area and slope for the 20 selected glaciers are shown in Fig. 5 along with the results from the manual digitization experiment and the flow line algorithm. Apart from a few cases (glaciers #6 & #14), length($_{area}$) shows unsystematic variability around the mean value, while length ($_{slope}$) is systematically lower than the mean (apart from glacier #15). Overall, the length values from the FLA are in most cases much closer (±5%) to the mean of the manual digitizations, confirming that it is worth the effort to apply it.

5.4 Length changes

Figure 9 shows examples of the length change assessment with the automatically created flow lines. Although in all cases the algorithm computed correct lines in regard to the glacier shapes and DEMs, the 'correct' length change values are a matter of debate. As changes in the shape of the terminus can have a strong impact on where terminus fluctuations can be measured, different methods of determination can lead to rather different results. For example, when the length change obtained from the flow lines (ΔL) is compared to the Euclidian distance between two terminus

positions (δL), we get (a good agreement) for the glacier in Fig. 9a ΔL = 1430 ± 101 m and δL = 1390 ± 101 m, but for the glacier in Fig. 9f a difference >100% (ΔL= 516 ± 101 m and δL= 1100 ± 101 m). Though the difference is obvious here, such a complete relocation of the flow line due to strong geometric changes of a glacier cannot easily be overcome and requires careful analysis. In principle, length changes of glaciers are terminus fluctuations rather than changes in the length of the flow line, so only the distance between the two terminus positions should be measured. This requires clearly defining where the terminus is, which is not an easy task when considering the multilobate glaciers in Fig. 9e and 9f. The same applies to the Blockade Glacier depicted in Fig. 6a. The decision which of the lobes should be used as the terminus can only be decided manually.

Fig. 9

A strong correlation (R^2= 0.88) between length change from the flow lines ΔL and Euclidian distance between the glacier terminus points δL is found. A more detailed analysis of this correlation was performed by plotting the differences of the two methods vs. length change ΔL (Fig. 10). For a retreat along a line that is curved rather than straight, δL should be smaller than ΔL. However, Fig. 10 shows that there are several exceptions from this rule (δL is larger than ΔL) which can only occur when something is wrong with at least one of the methods (e.g. a strong geometric change has moved the terminus point to a different location). In that sense the scatter plot presented in Fig. 10 allows the identification of glaciers that have to be checked before consideration. We found no correlation of the length change values with other topographic parameters (e.g. mean slope or mean elevation).

Fig. 10

6. DISCUSSION

6.1 Performance of the method

The FLA presented here creates a flow line and computes its length for several hundred glaciers in a fully automated way given that the two input datasets (glacier outline, DEM) are available. The main advantages compared to manual digitization come with the consistent, reproducible and fast computation. The computing time required to create the flow lines for an entire region depends on the number of large glaciers in the dataset. For small glaciers (< 5 km^2), lines are created within a second, while it may take 5 minutes or more for large glaciers (> 200 km^2). The examples in Fig. 9 illustrate that the performance of the FLA is in general very good (Fig. 9a-d), but can be quite different for the same glacier when a geometry change took place (Fig. 9e-f). Hence, visual inspection and manual corrections are still necessary to adjust poor line shapes or to select the most appropriate one. For the sample tested, 8% of lines had to be manually corrected to give glacier front changes, but to provide acceptable flow lines 17% needed correction. When the algorithm did not create a line, it is possible to adjust the distance between traverses (see Section 4.3c). Due to the conceptual idea behind the algorithm, flow lines can be deflected by a glacier tributary as the points are computed from segments that are located in the middle of every glacier "slice" (see white arrows on Fig. 9a and 9c). Though this effect changes the shape of the line slightly, it has no significant influence on the total length.

6.2 Comparison with manual approaches

Manual digitizing is considered as being the most accurate approach to create flow lines. However, the overlay in Fig. 4 and the direct comparison of length values in Fig. 7 reveal that this is not really

the case. For example, the digitized flow lines for a comparably simple shaped glacier depicted in Fig. 8b, reveals a rather high variability (between -7 and +19%) which is actually the largest for any of the 20 glaciers analysed here. The length values of the FLA are only in 3 out of 20 cases outside the range of the manual digitizations and in these cases the algorithm failed because of very complex glacier geometry or due to the limitations of the method. On the other hand, the outliers stress the importance of a having a visual inspection of all flow lines generated by the FLA and applying a manual correction where required. To obtain comparable results, it is mandatory to provide starting and end points of the flow line. The clear drawbacks of the manual digitizations are the time consuming work for a large number of glaciers, the inconsistency of the digitization, and the point that the work needs to be done again for other inventories (either past or future).

6.3 Computed length values and length changes

Comparing two glacier datasets from different sources is not straightforward, because the datasets could have been obtained from different techniques or methods (i.e. using remote sensing satellite data, airborne photographs, maps, automatic or manual digitizing) resulting in large potential mismatches. For example, even if the two datasets result from Landsat scenes, snow cover and cloud conditions could lead to differences in interpretation of the outline and hence to different flow lines. As many glaciers have experienced strong changes in their geometry in response to climate forcing during the past decades, it might be required to shift a former flow line to a different part of a glacier. In the case of a period dominated by glacier retreat, tributary glaciers might have been separated, or rock outcrops might have appeared somewhere on the glacier (e.g. Paul et al., 2007). Indeed, as the FLA is based on DEMs, a change in surface topography will also affect the resulting vector lines (e.g. when artefacts are present in the accumulation region of DEMs derived from optical sensors). However, the resulting changes in flow line position are small compared to changes resulting from changes of the outline.

7. CONCLUSIONS

This study presented a new and automated method called FLA to create glacier flow lines from only two input datasets (glacier outlines and DEM). The algorithm is based on open source software using the GDAL-OGR library and Python scripting. It uses a two dimensional geometry concept (glacier axis) to identify a series of centre points in each glacier that are connected to give the flow line. Difficult issues such as the requirement to not cross the outline or internal polygons (e.g. from rock outcrops) are solved by the algorithm. The highest points as computed from the DEM and the outlines require, however, visual control and possibly correction. This is particularly important when glacier outlines from two inventories are used to perform change assessment.

We have created about 400 glacier flow lines for two points in time using the outlines from the DLG and a new inventory. Length changes were calculated for 388 common glaciers, indicating a weak dependence of length changes on original glacier length. The Euclidian distance between the two terminus positions gives values that are too small in most cases, and strong deviations point to glaciers that require visual control. A manual digitizing experiment revealed good visual agreement with the FLA computed flow lines and a quantitative comparison of the length values showed that for most of the cases they are within the variability of the values derived from the manual digitization. Comparison with length values derived from glacier area or simple topographic modelling (elevation range and slope) revealed statistically much poorer results. However, some manual intervention remains (e.g. shifting starting points, editing obviously wrong flow lines) necessary to create a high quality product. In any case, we conclude that automated computation

with the FLA is very good alternative to a full manual digitization when it comes to large glacier samples and repeat application.

Acknowledgments
This study was performed in the framework of the ESA GlobGlacier project (21088/07/I-EC) and was partly also supported by the ESA Glaciers_cci project (4000101778/10/I-AM). Landsat scenes, DLG outlines, DRG maps and the DEM NED were obtained from the USGS (http://glovis.usgs.gov and http://seamless.usgs.gov). The ASTER GDEM is a product of Japan's Ministry of Economy, Trade and Industry (METI) and NASA and was downloaded from http://www.gdem.aster.ersdac.or.jp. We thank H. Frey for his help with the digitizing experiment and the two reviewers I. Evans and P. Leclercq for their very detailed and constructive comments.

REFERENCES

Arendt, A., et al., 2012, Randolph Glacier Inventory [v1.0]: A Dataset of Global Glacier Outlines. Global Land Ice Measurements from Space, Boulder Colorado, USA. Digital Media.

Bahr, D.B., 1997. Width and length scaling of glaciers. Journal of Glaciology, 43(145), 557-562.

Berthier, E., Schiefer, E., Clarke, G. K. C., Menounos, B. and F. Rémy, 2010. Contribution of Alaskan glaciers to sea-level rise derived from satellite imagery. Nature Geoscience, 3(2), 92-95.

Bodansky, E., Gribov, A. & Pilouk, M., 2002. Smoothing and compression of lines obtained by raster-to-vector conversion. Graphics Recognition Algorithms and Applications, 256-265.

Bolch, T., Menounos, B. & Wheate, R., 2010. Landsat-based inventory of glaciers in western Canada, 1985-2005. Remote Sensing of Environment, 114(1), 127-137.

Citterio, M., Paul, F., Ahlstrøm, A.P., Jepsen, H. F., Weidick, A., 2009. Remote sensing of glacier change in West Greenland: accounting for the occurrence of surge-type glaciers. Annals of Glaciology, 50(53), 70-80.

De Berg, M., Cheong, O. & Van Kreveld, M., 2008. Computational geometry: algorithms and applications, Springer-Verlag New York Inc.

Denton, G.H. & Field, W.O., 1975. Glaciers of the Alaska Range. Mountain Glaciers of the Northern Hemisphere. US Army Corps of Engineers, 2, 573–620.

Gesch, D.B., 2007. Chapter 4–The national elevation dataset. The DEM user's manual. Second edition. American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland, USA, 99–118.

Glasser, N.F., Harrison, S., Jansson, K. N., Anderson, K. and Cowley, A., 2011. Global sea-level contribution from the Patagonian Icefields since the Little Ice Age maximum. Nature Geoscience, 4(5), 303-307.

Haeberli, W. & Hoelzle, M., 1995. Application of inventory data for estimating characteristics of and regional climate-change effects on mountain glaciers: a pilot study with the European

Alps. Annals of Glaciology, 21, 206-212.

Hall, D.K., Bayr, K. J., Schöner, W., Bindschadler, R. A., Chien, J.Y.L., 2003. Consideration of the errors inherent in mapping historical glacier positions in Austria from the ground and space (1893-2001). Remote Sensing of Environment, 86(4), 566-577.

Hoelzle, M., Haeberli, W., Dischl, M., Peschke, W., 2003. Secular glacier mass balances derived from cumulative glacier length changes. Global and Planetary Change, 36(4), 295-306.

Klok, E.J. & Oerlemans, J., 2004. Climate reconstructions derived from global glacier length records. Arctic, Antarctic, and Alpine Research, 36(4), 575-583.

Le Bris, R., Paul, F., Frey, H., Bolch, T., 2011. A new satellite derived glacier inventory for Western Alaska. Annals of Glaciology, 52(59), 135-143.

Leclercq, P.W. & Oerlemans, J., 2011. Global and hemispheric temperature reconstruction from glacier length fluctuations. Climate Dynamics 38, pp 1065-1079 (2012).

Leclercq, P., J. Oerlemans and J. Cogley 2011. Estimating the Glacier Contribution to Sea-Level Rise for the Period 1800–2005. *Surveys in Geophysics*, 32(4): 519-535.

Lemke, P. et al., 2007. Observations: Changes in snow, ice and frozen ground. Title: Climate change, 337-383.

Lüthi, M.P., Bauder, A. & Funk, M., 2010. Volume change reconstruction of Swiss glaciers from length change data. Journal of Geophysical Research, 115, p.F04022.

Manolopoulos, Y., Nanopoulos, A., Papadopoulos, A. N. and Theodoridis, Y., 2005. R-Trees: Theory and Applications, Advanced Information and Knowledge Processing Series, Springer. 195pp

Moore, A.W., 1991. An intoductory tutorial on kd-trees. Computer Laboratory, University of Cambridge. 20pp

Müller, F., Caflisch, T. and Müller, G. (eds.), 1977. Instructions for the compilation and assemblage of data for a world glacier inventory. IAHS(ICSI)/UNESCO report, Temporal Technical Secretariat for the World Glacier Inventory (TTS/WGI), ETH Zurich, Switzerland.

Nussbaumer, S.U. & Zumbühl, H.J., 2011. The Little Ice Age history of the Glacier des Bossons (Mont Blanc massif, France): a new high-resolution glacier length curve based on historical documents. Climatic Change, 1-34.

Oerlemans, J., 2001: Glaciers and Climate Change. A.A. Balkema Publishers, 148 pp. ISBN 9026518137.

Oerlemans, J., 2005. Extracting a Climate Signal from 169 Glacier Records. Science, 308(5722), 675-677.

Oerlemans, J., M. Dyurgerov, and R. S. W. van de Wal, 2007, Reconstructing the glacier contribution
to sea-level rise back to 1850, The Cryosphere, 1(1), 59-65.

Oerlemans, J., 2008. Minimal Glacier Models. Igitur, Utrecht Publishing & Archiving Services, Universiteitsbibliotheek Utrecht, 90pp.

Paul, F., Kääb, A. & Haeberli, W., 2007. Recent glacier changes in the Alps observed by satellite: consequences for future monitoring strategies. Global and Planetary Change, 56(1-2), 111-122.

Paul, F. et al., 2009. Recommendations for the compilation of glacier inventory data from digital sources. Annals of Glaciology, 50(53), 119-126.

Racoviteanu, A.E., Paul, F., Raup, B., Khalsa, S. J. S., Armstrong, R., 2009. Challenges and recommendations in mapping of glacier parameters from space: results of the 2008 Global Land Ice Measurements from Space (GLIMS) workshop, Boulder, Colorado, USA. Annals of Glaciology, 50(53), 53-69.

Raup, B. et al., 2007. Remote sensing and GIS technology in the Global Land Ice Measurements from Space (GLIMS) project. Computers & Geosciences, 33(1), 104-125.

Tsaparas, P. & Science, U. of T.D. of C., 1999. Nearest Neighbor Search in Multidimensional Spaces: Depth Oral Report, University of Toronto, Dept. of Computer Science, 53pp.

WGMS, 2008. Global glacier changes: facts and figures. Geneva: UNEP/WGMS, 45pp.

Williams Jr, R.S., Hall, D. K., Sigurdsson, O., Chien, J. Y. L., et al., 1997. Comparison of satellite-derived with ground-based measurements of the fluctuations of the margins of Vatnajökull, Iceland, 1973-1992. Annals of Glaciology, 24, 72-80.

Lopez, P., Chevallier, P., Favier, V., Pouyaud, B., Ordenes, F. and Oerlemans, J., 2010. A regional view of fluctuations in glacier length in southern South America. Global and Planetary Change, 71, 85-108.

Quinn, P., K. Beven, P. Chevallier and O. Planchon, (1991): The prediction of hillslope flow paths for distributed hydrological modeling using digital terrain models. Hydrological Processes, 5 (1), 59-79.

FIGURE CAPTIONS

Fig. 1. Overview of the study region in Alaska. The thick blue lines mark glacier extents in the Tordrillo and Chigmit Mountains. Black dots indicate the twenty glaciers used for comparison with manual digitizing (see text for explanation). The inset shows the location of the study region in Alaska.

Fig. 2. Illustration of the principle work flow of the algorithm describing its major steps. Letters on the left top corner of each illustration refer to the respective step described on the flow chart. Blue and red squares in A-B represent enlargements showed in E and F and in G.

Fig. 3. Flow lines (dashed) as computed with the FLA for the two outlines (solid) of the DLG extent (red) and the new inventory (blue) for Wolverine and Gulkana glaciers. Red and yellow dots are the starting and terminus points.

Fig. 4. Comparison of the manually digitized flow lines (thin, grey) and the automatically derived lines (thick, black) after smoothing. Rasterized lines are the glacier outlines. Red and yellow dots are the starting and terminus points, respectively.

Fig. 5. Relative length differences of the manually digitized (grey) and automatically created flow lines for 20 selected glaciers. Y axis represents the relative mean difference of the automated lengths compared to the manual values. Horizontal red bars show the length of each glacier computed with the FLA, while black and blue bars show the length estimated from area and slope, respectively. Glaciers (ID) are ordered according to the glacier length (as derived from the algorithm).

Fig. 6. Illustration of poor results obtained by the algorithm. A, B and C correspond to glaciers #8 (Blockade Glacier), #15 and #16 (two unnamed glaciers), respectively.

Fig. 7. (a) Log-log plot of glacier length vs. glacier area for a subset of the new Alaskan glacier inventory. (b) Relative differences of length as estimated from glacier area ($Length_{area}$) and length estimated from the algorithm ($Length_{FLA}$) vs. $Length_{FLA}$.

Fig. 8. (a) Glacier length as derived from elevation range and mean slope ($Length_{slope}$) vs. $Length_{FLA}$. (b) Relative differences of $Length_{slope}$ and $Length_{FLA}$ vs. $Length_{FLA}$.

Fig. 9. Comparison of length change assessments. Images from a) to d) show examples with good agreement, whereas e) and f) are cases providing wrong results. Dotted lines in black refer to flow lines computed from the new Alaskan glacier inventory, and yellow lines (also dotted) are computed from the DLG outlines. White arrows in a) and c) indicate where glacier tributaries influence the shape of the lines. False colour composite images (TM bands 543 as RGB) are displayed in the background.

Fig. 10. A scatter plot illustrating the difference between the Euclidian distance of the two terminus position and their distance as derived from the respective segment of the FLA vs. length change.

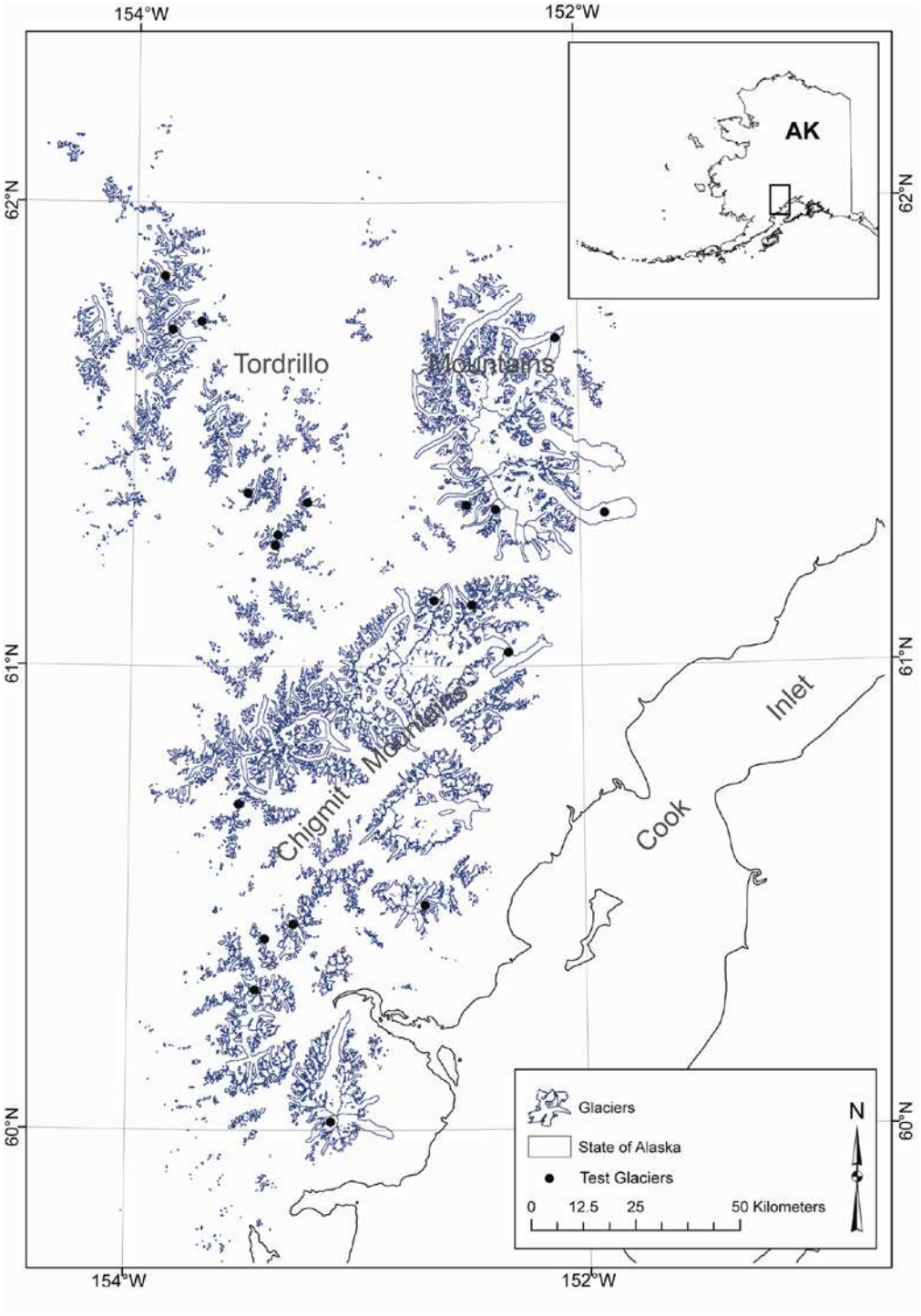
154°W
152°W
62°N
61°N
60°N
AK
Tordrillo
Mountains
Chigmit
Mountains
Cook
Inlet
Glaciers
State of Alaska
Test Glaciers
0
12.5
25
50 Kilometers
N

Fig. 1

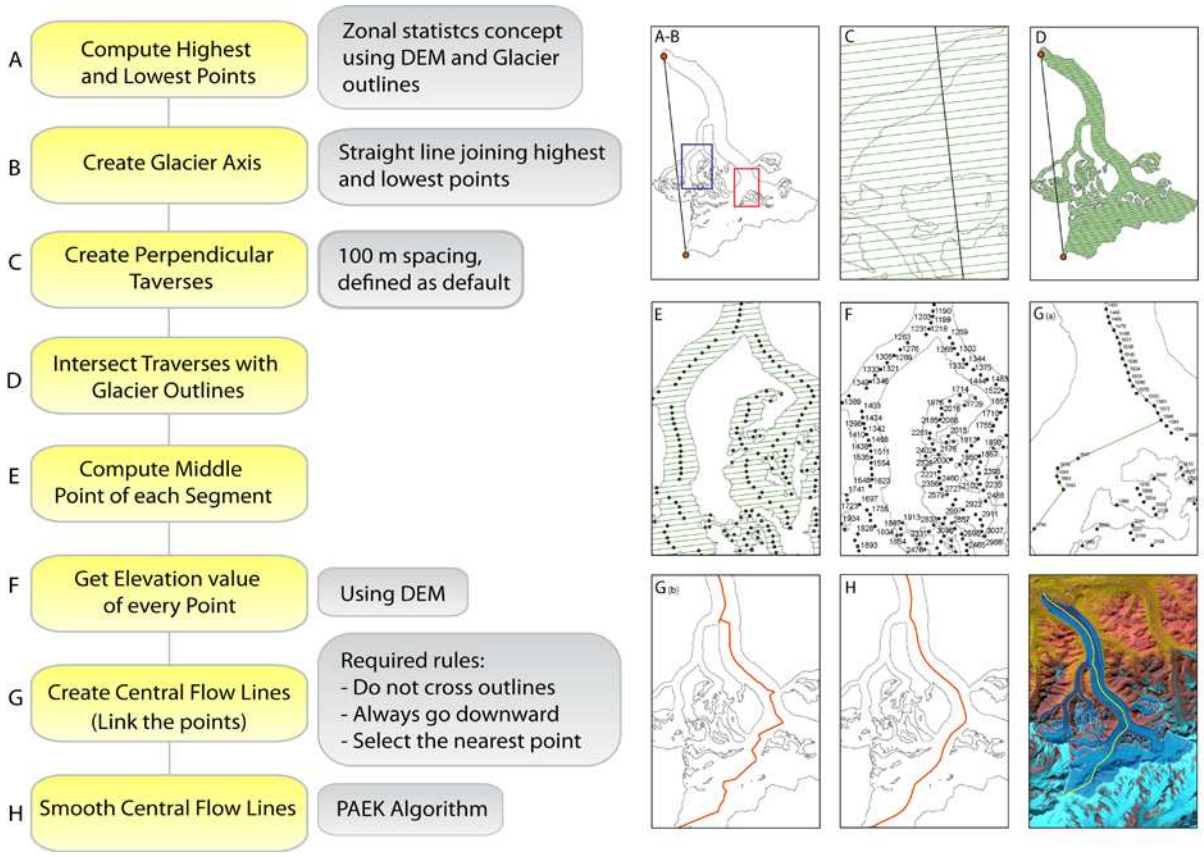
A
Compute Highest and Lowest Points
Zonal statistcs concept using DEM and Glacier outlines
B
Create Glacier Axis
Straight line joining highest and lowest points
C
Create Perpendicular Taverses
100 m spacing, defined as default
D
Intersect Traverses with Glacier Outlines
E
Compute Middle Point of each Segment
F
Get Elevation value of every Point
Using DEM
G
Create Central Flow Lines (Link the points)
Required rules:
- Do not cross outlines
- Always go downward
- Select the nearest point
H
Smooth Central Flow Lines
PAEK Algorithm
A-B
C
D
E
F
G (a)
G (b)
H

Fig. 2

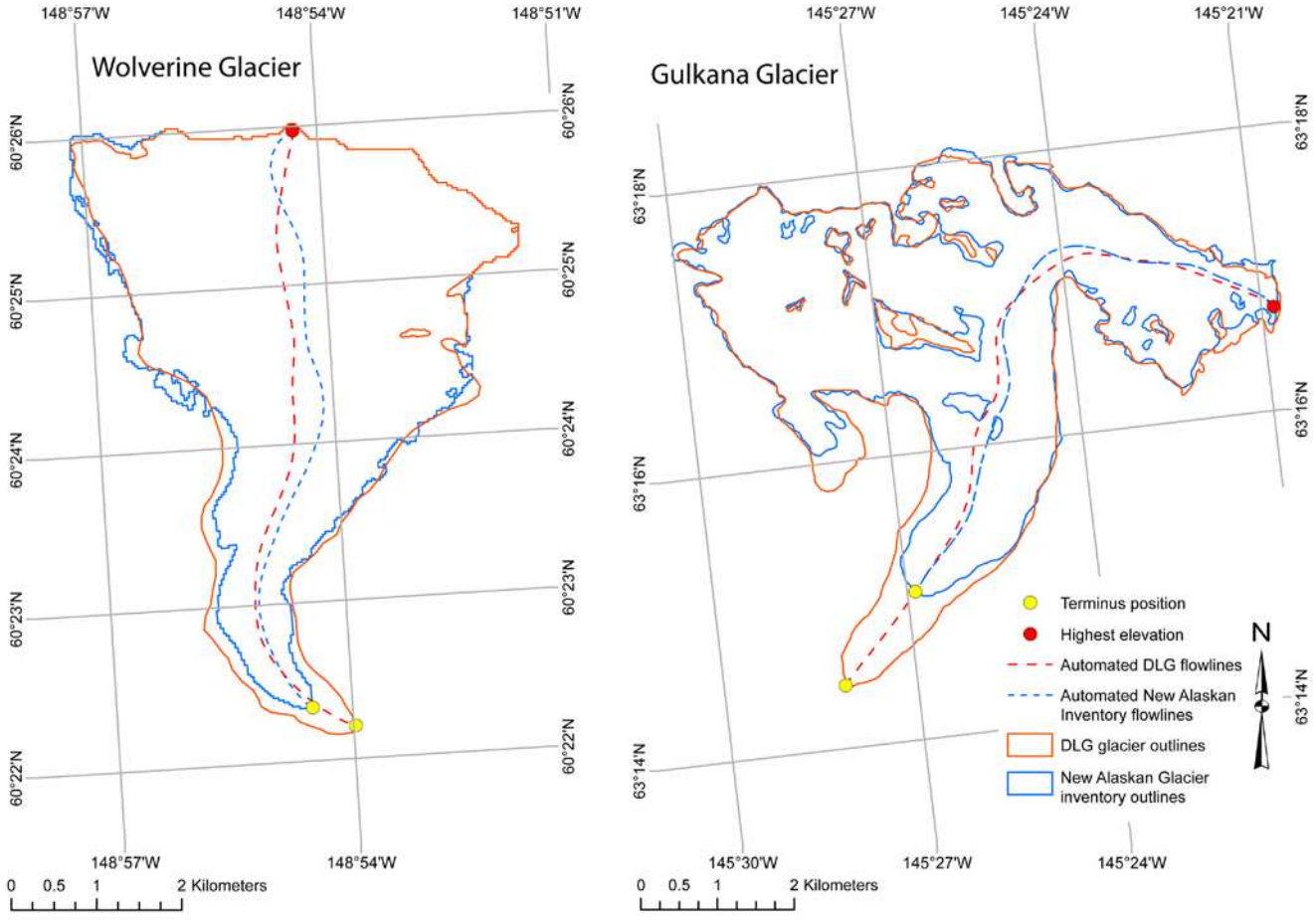
Wolverine Glacier
148°57'W
148°54'W
148°51'W
60°26'N
60°25'N
60°24'N
60°23'N
60°22'N
0 0.5 1 2 Kilometers
Gulkana Glacier
145°27'W
145°24'W
145°21'W
63°18'N
63°16'N
63°14'N
145°30'W
Terminus position
Highest elevation
Automated DLG flowlines
Automated New Alaskan Inventory flowlines
DLG glacier outlines
New Alaskan Glacier inventory outlines
N

Fig. 3

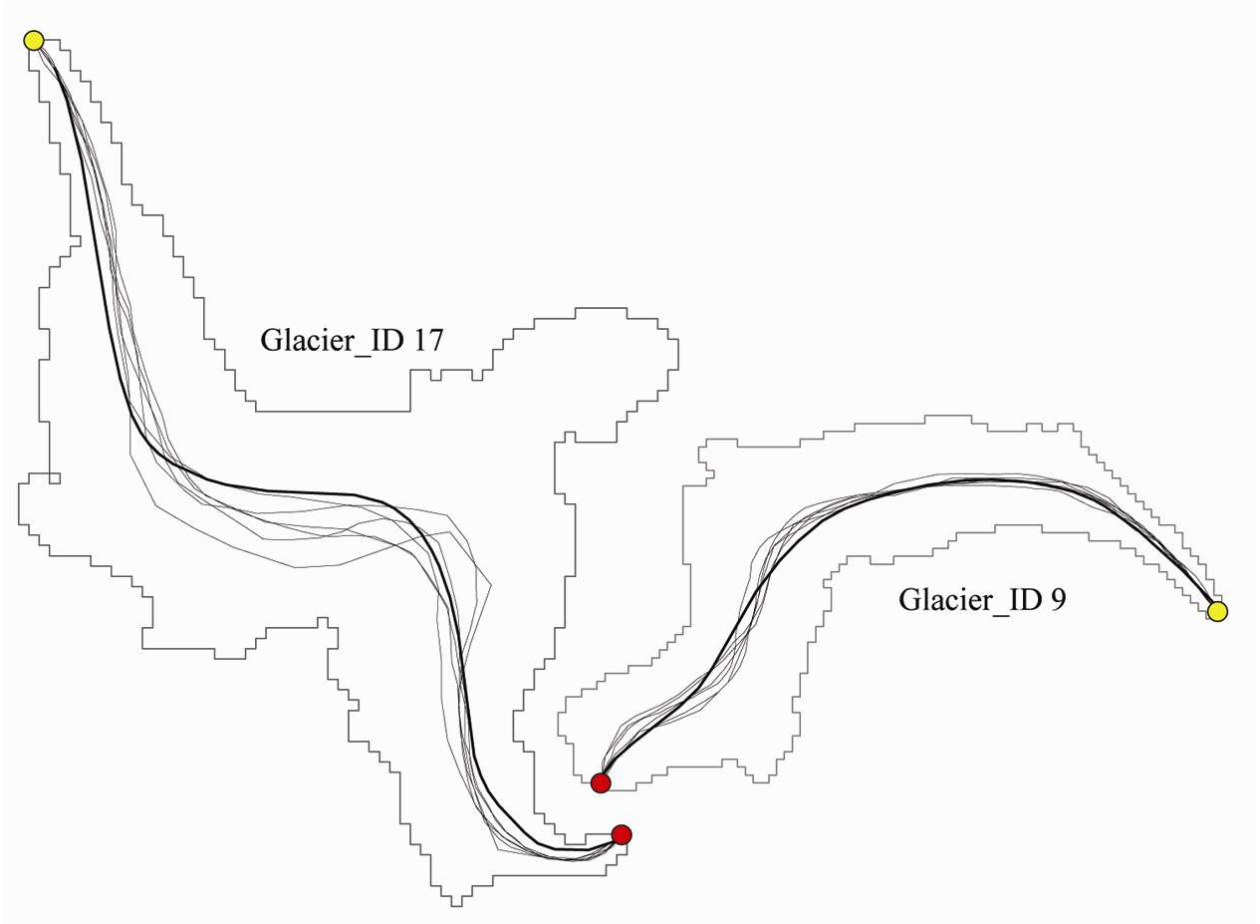
Glacier_ID 17
Glacier_ID 9

Fig. 4

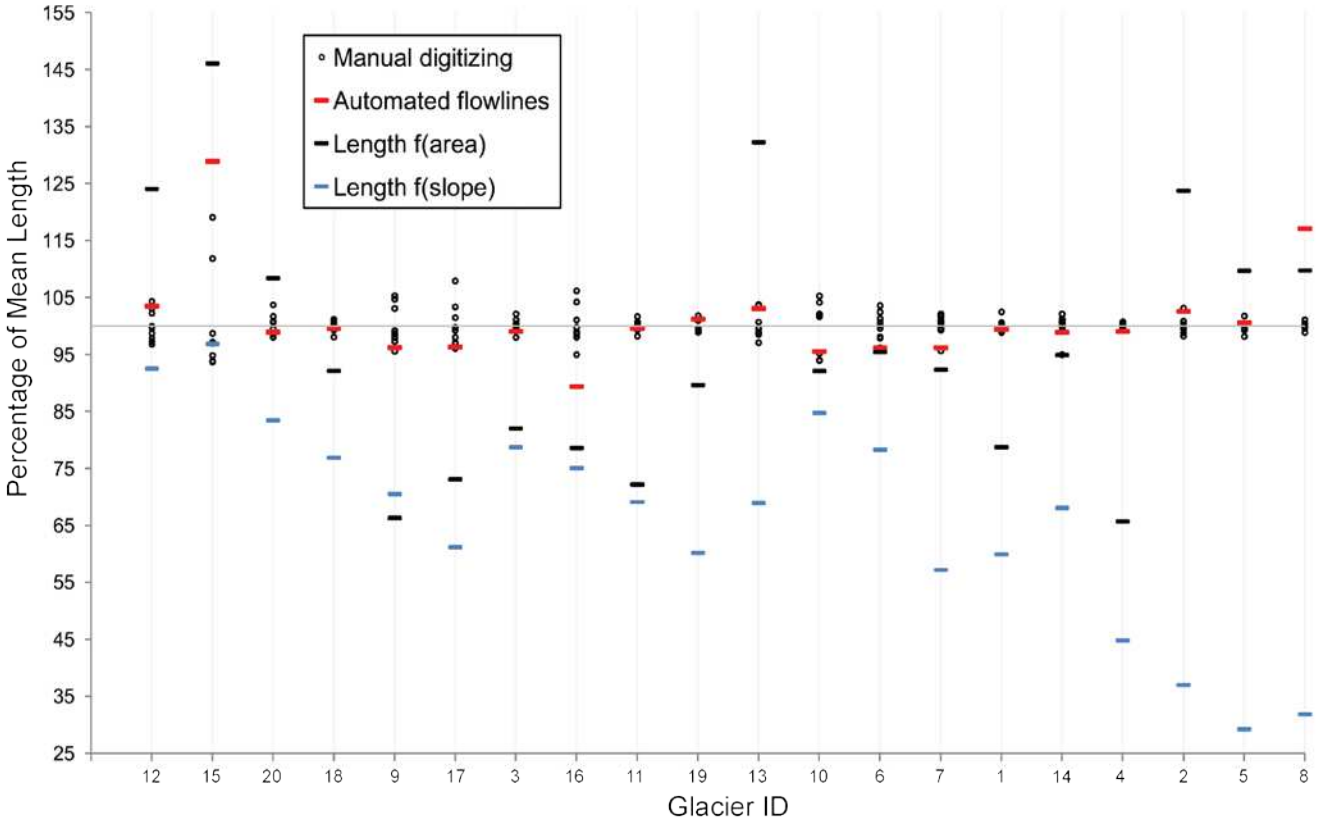
Manual digitizing
Automated flowlines
Length f(area)
Length f(slope)
Percentage of Mean Length
155
145
135
125
115
105
95
85
75
65
55
45
35
25
12
15
20
18
9
17
3
16
11
19
13
10
6
7
1
14
4
2
5
8
Glacier ID

Fig. 5

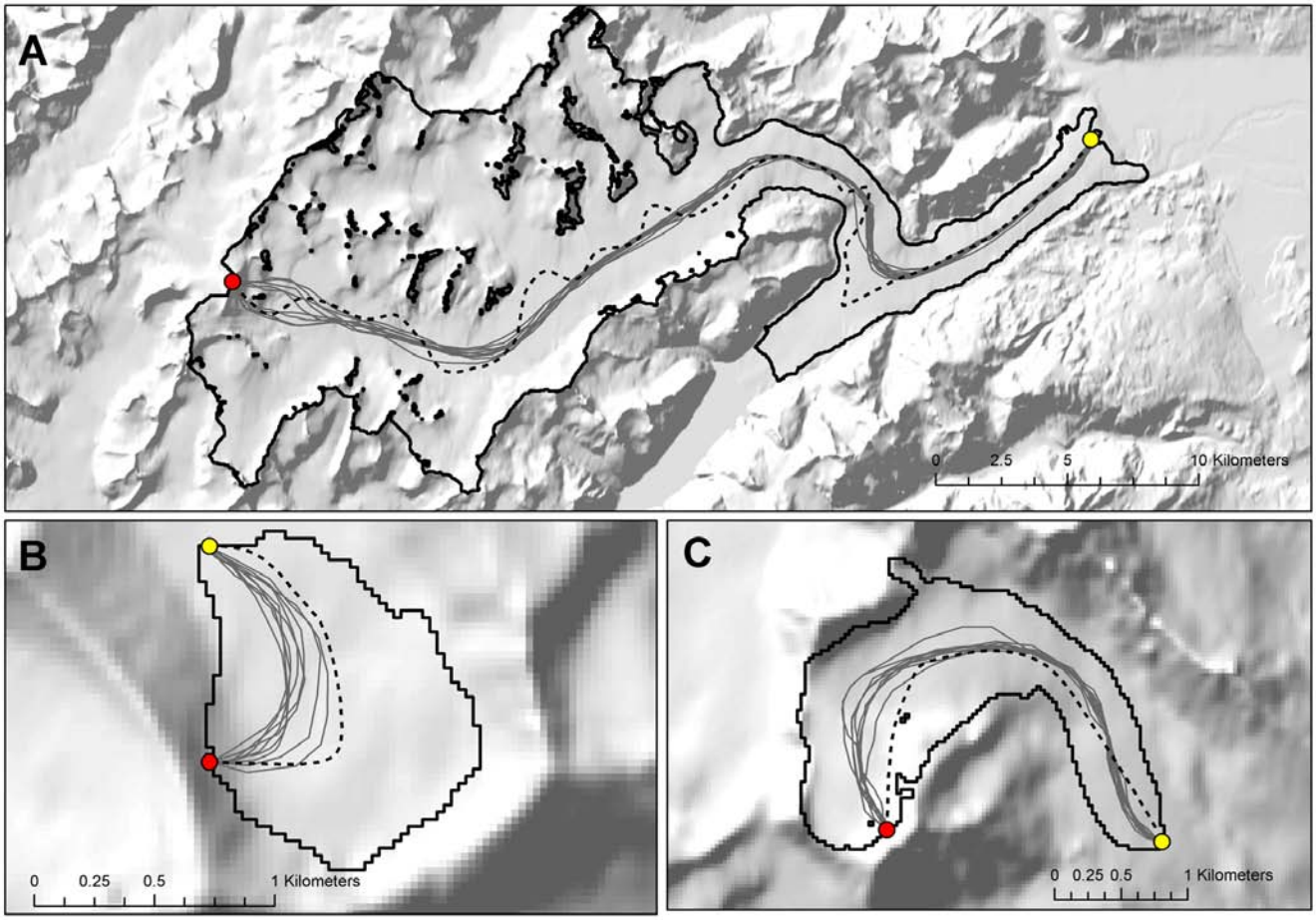
A
0 2.5 5 10 Kilometers
B
0 0.25 0.5 1 Kilometers
C
0 0.25 0.5 1 Kilometers

Fig. 6

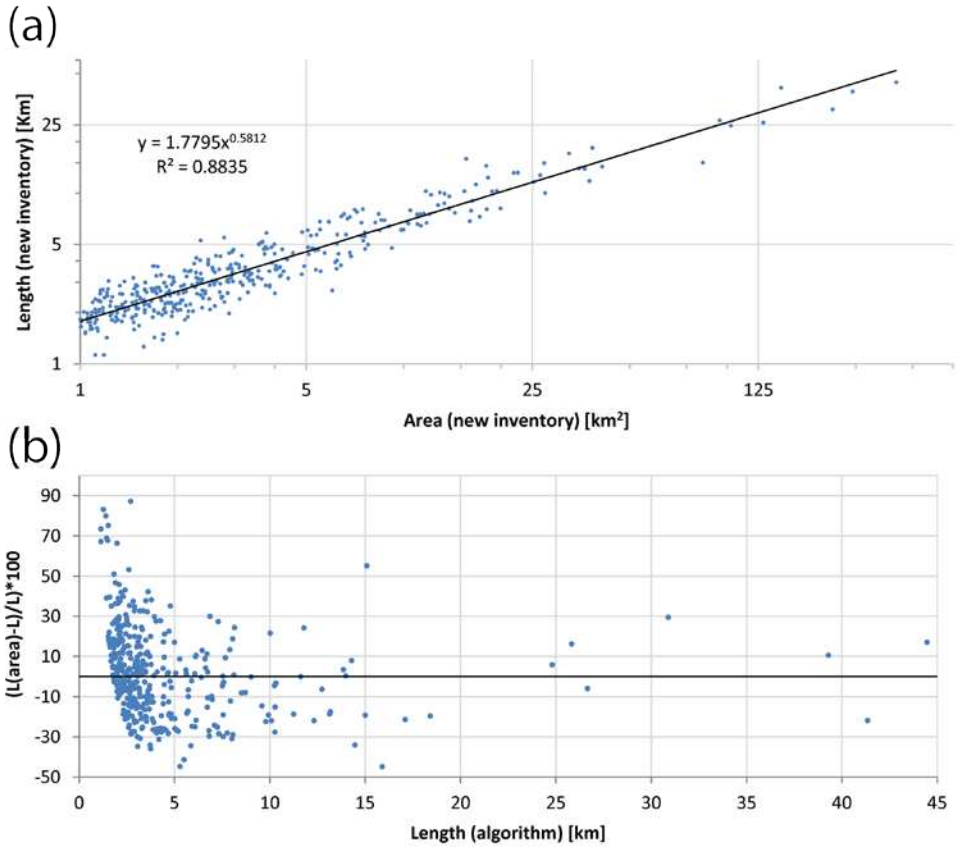
(a)
y = 1.7795x^0.5812
R² = 0.8835
Length (new inventory) [Km]
Area (new inventory) [km²]
(b)
(L(area)-L)/L)*100
Length (algorithm) [km]

Fig. 7

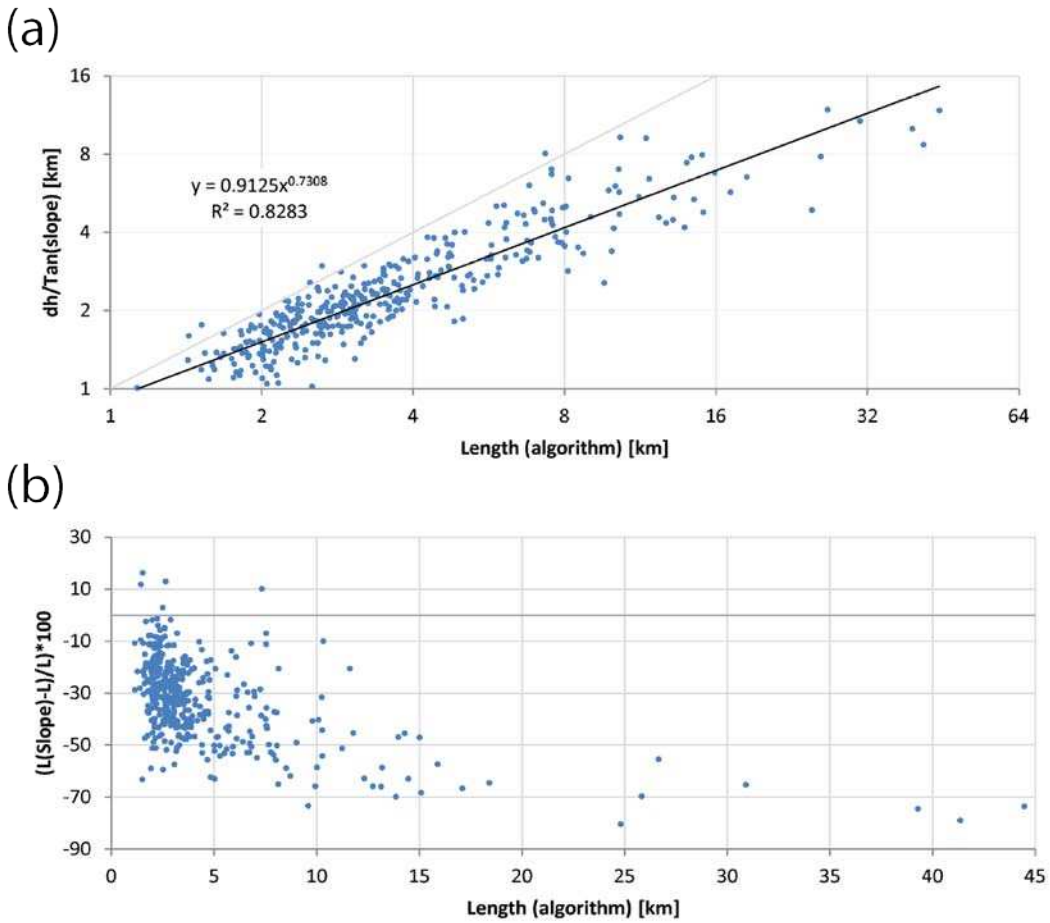
(a)
y = 0.9125x^0.7308
R² = 0.8283
dh/Tan(slope) [km]
Length (algorithm) [km]
(b)
(L(Slope)-L)/L)*100
Length (algorithm) [km]

Fig. 8

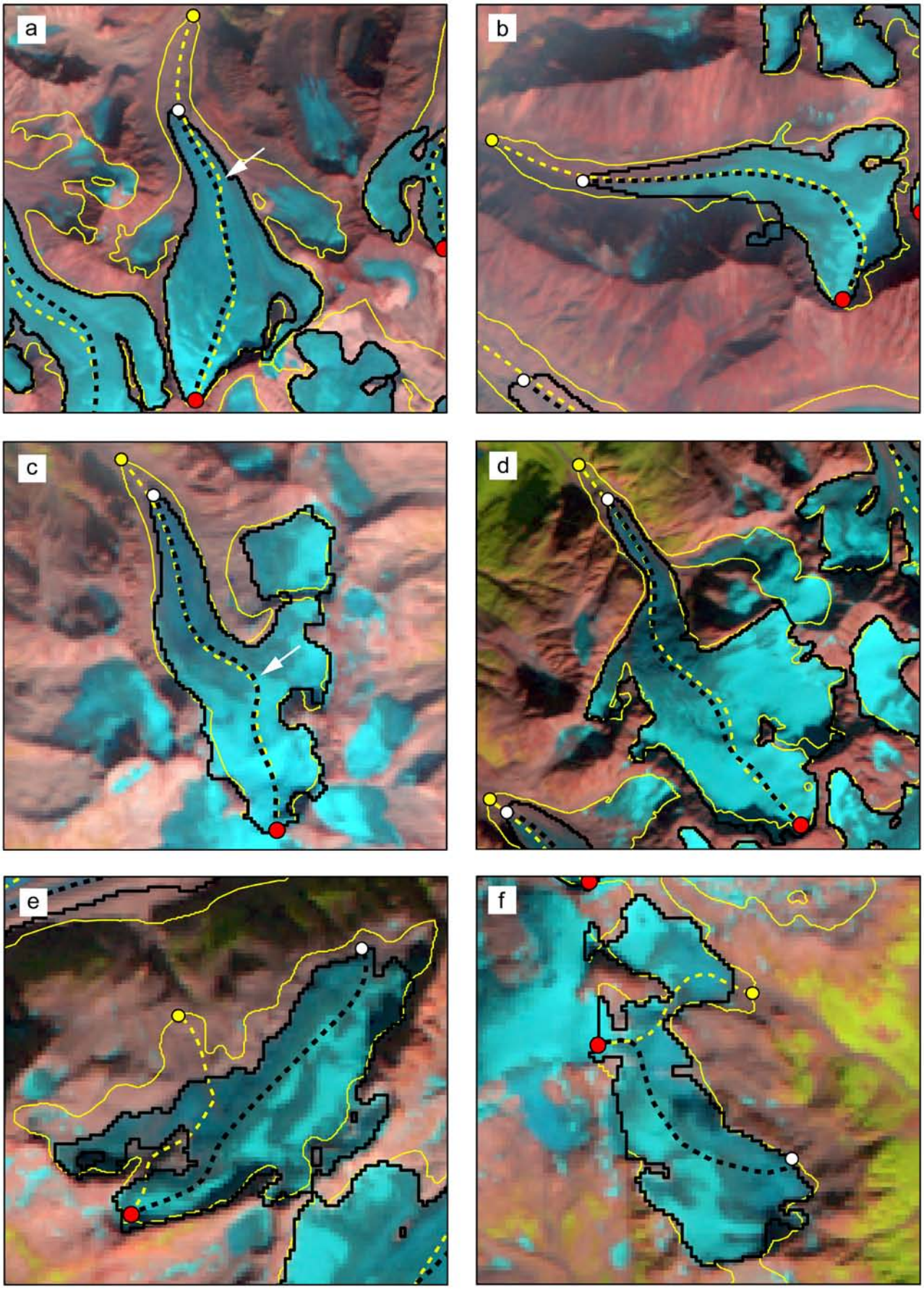

Fig. 9

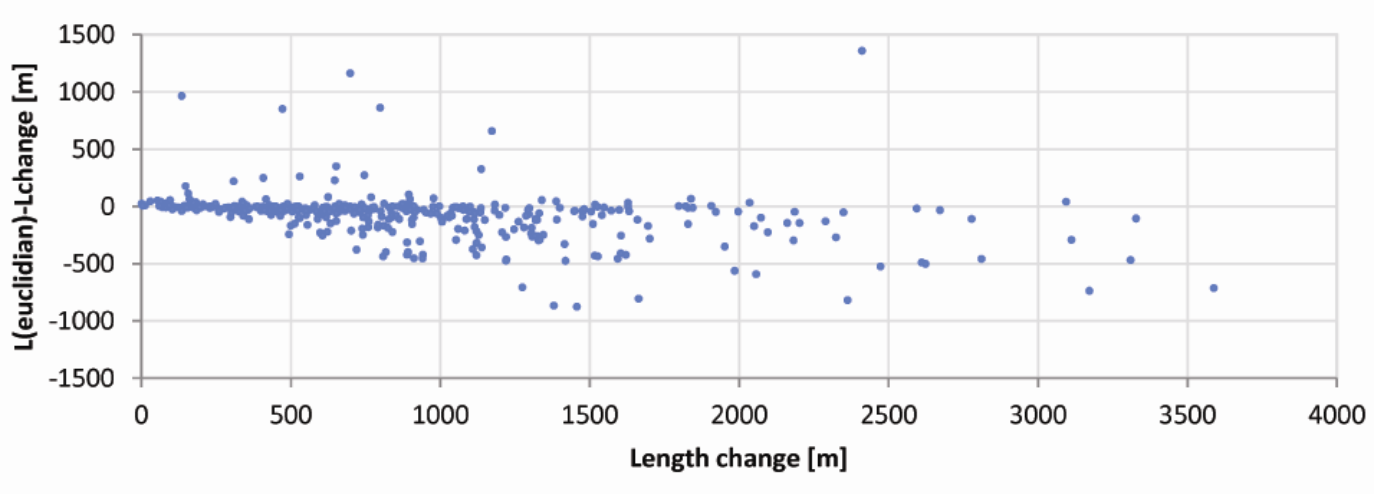
L(euclidian)-Lchange [m]
1500
1000
500
0
-500
-1000
-1500
0
500
1000
1500
2000
2500
3000
3500
4000
Length change [m]

Fig. 10